DIFFUSE MODELS FOR SAMPLING

by

David A. Lane and William D. Sudderth

Technical Report No. 287

June 1977

ABSTRACT

As a natural, intuitive model for inferences about certain characteristics of finite populations, Bruce Hill has proposed a sequence of exchangeable variables $X_1,\ldots,X_{n+1}$ which have distinct values with probability one and have the property that, conditional on $X_1,\ldots,X_n$, the next observation X_{n+1} is equally likely to fall in any of the $n+1$ intervals determined by $X_1,\ldots,X_n$. Harold Jeffreys had previously assumed such a model (in the case $n = 2$) for normal measurements with unknown mean and variance. Hill has shown that, for $n \geq 1$, there exist no countably additive distributions with the prescribed properties. It is shown here that finitely additive distributions with these properties do exist for all n and have a number of interesting properties.

AMS 1970 subject classifications. 62A15, 28A35.

Key Words and Phrases. Sampling models, diffuse priors, finite additivity, exchangeable variables, product measures.

1. Introduction

Bruce Hill (1968) described a sampling situation in which the numerical characteristic under observation has an arbitrary or "rubbery" scale, and prior information about the population distribution on this scale (assumed to be continuous) is vague. In this situation, Hill argued, the numerical values of a sample convey negligible information about the overall population values, although they do induce what might be called a predictive ordering on those values. For example, it is reasonable to assume that the second observation is equally likely to be bigger or smaller than the first observation, whether the first is 5 or 50. Here is a way to formulate this property mathematically: let $(X_1, \ldots, X_n)$ represent a sample drawn without replacement from the population; let $J_1, \ldots, J_{n+1}$ denote the $n+1$ (random) intervals into which $(X_1, \ldots, X_n)$ partitions R; and let X_{n+1} represent a further sample (still without replacement) from the population. Then the distribution for X_{n+1} should assign equal weight to each J_i, regardless of the numerical values of $X_1, \ldots, X_n$. More precisely,

(a) $P[(X_1, \ldots, X_n) \in A \text{ and } X_{n+1} \in J_i] = 1/n+1\ P((X_1, \ldots, X_n) \in A)$, for $1 \le i \le n+1$ and every Borel set $A \subseteq R^n$.

Clearly, if a <u>fixed</u> population is sampled, (a) cannot be achieved. What is sought is a prior distribution on populations, such that if "P" represents unconditional probability with respect to this prior, (a) is true. Such a prior is noninformative in the sense that no matter what values are observed for the first n individuals sampled, the $n+1^{st}$ individual is equally likely to fall between any two of them.

Hill posed the question of the existence of such priors in a broader setting: for fixed n, he asked for the construction of an exchangeable sequence of random variables $X_1, \ldots, X_{n+1}$ and a probability measure P such that (a) is true. He proved that, for all $n \ge 1$, no solution exists to this problem if P is required to be countably additive. He left the problem open in case P is only required to be finitely additive. Section 2 of this paper contains a formal statement of this problem.

For $n = 2$, Hill's problem has a solution in a setting quite apart from finite populations. This solution involves modifying a construction due to Jeffreys (1932) dealing with normal measurement error when the mean μ and variance σ^2 are unknown. Again, the problem involves noninformative priors: Jeffreys computed a prior on the parameter space by assuming (a) (which he considered an obvious fact in this case) and showing that only the improper prior, $\sigma^{-2}\, d\mu d\sigma^2$, is consistent with (a). If the parameter space is equipped with Jeffreys' prior, and X_1, X_2, and X_3 are assumed to be conditionally independent $N(\mu, \sigma^2)$, then a sampling model satisfying (a) is obtained. Substituting a natural finitely additive prior for Jeffreys' prior yields a solution to Hill's problem for $n = 2$; this is carried out in section 3 below.

In section 5, we return to the context of sampling from a finite population and solve Hill's problem in this case. In this section it is assumed that the population size, m, is known. A class of finitely additive measures on R^m, called <u>strategic product measures</u>, is defined. These measures play the role of prior distributions on the space of finite populations of size m. Let ν^m be a strategic product measure on R^m; another measure on R^m, denoted $\mathcal{P}(\nu^m)$ is defined in section 5. Under $\mathcal{P}(\nu^m)$, the m coordinate functions play the role of sampling variables from a population chosen by ν^m. Theorem 2--together with Lemma 4.2--characterize the class of strategic product measures ν^m such that, under $\mathcal{P}(\nu^m)$, the coordinate functions satisfy (a) for all $n \le m-1$. In particular, Theorem 2 shows that solutions to Hill's problem exist for all finite n.

If the population size is unknown, the prior must live on $\bigcup_{m=1}^{\infty} R^m$. It is easy to construct such priors. Just "pick an integer N at random" according to some purely finitely additive probability on the positive integers. Then construct a strategic product measure ν^N on R^N (verifying the conditions of Theorem 2) and its associated sampling distribution $\mathcal{P}(\nu^N)$. Since N is finite but larger than any given integer with probability one, this construction will yield an infinite sequence of exchangeable variables verifying (a) for every n. A somewhat more formal solution to this problem is presented in section 6.

Sections 7-9 discuss various properties of arbitrary solutions to Hill's problem, and section 10 presents some comments on unresolved questions connected with ideas in this paper.

The material in sections 4, 5, and 8 indicates that strategic product measures with widely different properties lead to sampling variables satisfying (a). (For example, see Theorem 6 and its corollary). As Hill's work shows, inference on certain population characteristics (including percentiles) depends on the prior only through (a). In this sense, all the priors satisfying (a) are noninformative. Further work may reveal inference questions with respect to which some of the priors constructed here are appropriate and some are not.

The fact that finitely additive priors which lead to sampling variables satisfying (a) exist has an important consequence for Hill's theory: Heath and Sudderth (1976) show that inferences are coherent when (and only when) they are consistent with a finitely additive prior.

2. The Problem.

Let R^n be real n-dimensional space and S_n the set of all permutations on $\{1, \ldots, n\}$. Each π in S_n induces a transformation on R^n, also denoted π by

$$\pi(x_1, \ldots, x_n) = (x_{\pi(1)}, \ldots, x_{\pi(n)}).$$

A probability on R^n is a finitely additive probability defined on all subsets of R^n. Let H_n be the collection of all probabilities β on R^n which satisfy the following conditions:

(i) Exchangeability: $\beta(A) = \beta(\pi(A))$ for every Borel set $A \subseteq R^n$ and $\pi \in S_n$.

(ii) No ties: $\beta\{x \in R^n\colon x_i = x_j\} = 0$ for $1 \leq i < j \leq n$.

(iii) For $1 \leq i \leq n$, the event that the nth coordinate function has rank i is independent of the first $n-1$ coordinate functions and has probability $1/n$. That is, for every Borel $A \subseteq R^{n-1}$ and $1 \leq i \leq n$,

$$\beta\{x \in R^n\colon (x_1, \ldots, x_{n-1}) \in A \text{ and } x_n = x_{(i)}\} = \frac{1}{n}\beta(A \times R), \tag{2.1}$$

where $x_{(i)}$ is the i^{th} smallest coordinate of $(x_1, \ldots, x_n)$.

Hill's problem, as discussed in section 1, is to show that H_n is nonempty for $n \geq 2$.

3. An example: Three Normal Measurements with Mean and Variance Unknown

For $n = 2$, a modification of the argument in Jeffreys (1932) yields a solution to Hill's problem. Identify the normal parameter space $\{(\mu, \sigma^2): \mu \in R, \sigma^2 > 0\}$ with the affine group in one dimension by associating (μ, σ^2) with the affine transformation $x \to \sigma x + \mu$. To model lack of knowledge about the true parameters, it is natural to take as prior measure on the parameter space an invariant measure on this group. So let Π be a finitely additive, left-invariant probability measure on the one-dimensional affine group (see Greenleaf (1969), p. 68), and let β be any probability on R^3 such that, if A is a Borel set, then

$$\beta(A) = \int N(A|\mu,\sigma^2)\, d\Pi(\mu,\sigma^2),$$

where $N(A|\mu,\sigma^2) = \iiint_A (2\pi\sigma^2)^{-3/2}\exp(-\frac{1}{2\sigma^2} \sum(x_i - \mu)^2)dx_1dx_2dx_3$.

That is, given (μ,σ^2), the coordinate functions--which we shall write as random variables X_1, X_2, and X_3--are independent normal with mean μ and variance σ^2.

Let $d = \frac{1}{2} \cdot |X_1 - X_2|$ and $m = \frac{1}{2}(X_1 + X_2)$. As Heath and Sudderth (1976) show, the posterior distribution of (μ, σ^2) given X_1 and X_2 is the same as the distribution of $(dKS^{-1} + m, 2d^2S^{-2})$ where K and S are independent standard normal variables. It then follows--just as in the countably additive case--that we may calculate the distribution of X_3 given X_1 and X_2 (for details, see section 9). If we denote this conditional distribution $\beta^{\cdot}_{x_1,x_2}$ then, for $A \subset R$,

$$(3.1) \quad \beta_{x_1,x_2}(A) = \iint_A (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)\,dx\,d(\mu,\sigma^2|x_1,x_2)$$

where $d(\mu,\sigma^2|x_1,x_2)$ is the posterior distribution for (μ,σ^2) given $X_1 = x_1$ and $X_2 = x_2$.

In particular, if Φ represents the standard normal c.d.f., and f is the standard normal density (and for convenience, suppose $x_1 < x_2$), then

$$\beta_{x_1,x_2}\{x: x_1 \le x \le x_2\} = \int[\Phi(\frac{x_2-\mu}{\sigma}) - \Phi(\frac{x_1-\mu}{\sigma})]\,d(\mu,\sigma^2|x_1,x_2)$$

$$= \int[\Phi\{(\sqrt{2}\,d)^{-1}s(x_2 - dkx^{-1} - m)\} - \Phi\{(\sqrt{2}\,d)^{-1}s(x_1 - dks^{-1} - m)\}f(k)f(s)dkds$$

$$= \int[\Phi\{(\sqrt{2}\,d)^{-1}sd(1 - ks^{-1})\} - \Phi\{(\sqrt{2}\,d)^{-1}sd(-1-ks^{-1})\}\,f(k)f(s)dkds$$

$$= \int[\Phi\{(2^{-\frac{1}{2}}(s - k)\} - \Phi(2^{-\frac{1}{2}}(-s-k)]f(k)f(s)dkds.$$

The last line is free of d and m and hence does not depend on the particular values of x_1 and x_2. Since the same expression is obtained if $x_1 \ge x_2$, $\beta_{x_1,x_2}\{x: x_1 \le x \le x_2\}$ is a constant function of the argument (x_1, x_2).

Moreover,

$$1/3 = \beta\{(x_1, x_2, x_3): x_1 \le x_3 \le x_2 \text{ or } x_2 \le x_3 \le x_1\}$$

$$= \int_{\{x_1<x_2\}}\beta_{x_1,x_2}\{x: x_1 \le x \le x_2\}\,d\gamma(x_1,x_2) + \int_{\{x_2<x_1\}}\beta_{x_1,x_2}\{x: x_2 \le x \le x_1\}d\gamma(x_1,x_2)$$

where γ is the marginal distribution of the first two coordinates. By the preceding paragraph, the integrand is a constant--hence the constant must be 1/3. Similarly, it can be shown that

$$\begin{aligned} 1/3 &= \beta_{x_1,x_2}\{x\colon x \le \min(x_1,x_2)\} \\ &= \beta_{x_1,x_2}\{x\colon \min(x_1,x_2) \le x \le \max(x_1,x_2)\} \\ &= \beta_{x_1,x_2}\{x\colon x \ge \max(x_1,x_2)\} \quad \text{a.s.} \quad \gamma. \end{aligned}$$

Thus β satisfies condition (iii) of section 2 with $n = 2$; since β clearly satisfies (i) and (ii), β is an element of H_2.

4. An alternative formulation of the problem.

There is a natural correspondence between H_n and the set O_n of probabilities α on R^n which satisfy these two conditions:

(iv) $\alpha\{x\colon x_1 < x_2 < \ldots < x_n\} = 1$,

(v) for $1 \le m < n$, the α-marginal distribution of any set of m coordinates (in ascending order) is the same.

To describe the correspondence, some notation is needed. For $x = (x_1, \ldots, x_n)$ in R^n, let $\operatorname{ord} x = (x_{(1)}, \ldots, x_{(n)})$ denote the vector whose coordinates are defined by: $\{x_{(1)}, \ldots, x_{(n)}\} = \{x_1, \ldots, x_n\}$ and $x_{(1)} \le \ldots \le x_{(n)}$. Also, to each probability α on R^n, associate its symmetrization $\beta = \mathcal{P}(\alpha)$, where

(4.1) $$\mathcal{P}(\alpha)(A) = \frac{1}{n!} \sum_{\pi \in S_n} \alpha\{x\colon \pi(x) \in A\},$$

for each $A \subseteq R^n$. Intuitively, the $\mathcal{P}(\alpha)$-distribution is obtained by choosing $x \in R^n$ according to α and then putting the coordinates in a random order.

The theorem below demonstrates that Hill's problem is equivalent to showing that O_n is nonempty for $n \ge 2$.

<u>Theorem 1</u>. <u>If</u> $\alpha \in O_n$, <u>then</u> $\mathcal{P}(\alpha) \in H_n$. <u>Conversely</u>, <u>if</u> $\beta \in H_n$ <u>and</u> α <u>is the distribution under</u> β <u>of</u> ord x, <u>then</u> $\alpha \in O_n$ <u>and</u> $\mathcal{P}(\alpha) = \beta$.

The proof of Theorem 1 will take up most of the remainder of this section. To prove the first assertion, fix $\alpha \in O_n$ and let $\beta = \mathcal{P}(\alpha)$. Clearly, β satisfies conditions (i) and (ii) in the definition of H_n. To verify (iii), assume $n > 1$ (for otherwise (iii) is trivial) and define a mapping $\pi \to \bar{\pi}$ from S_n to S_{n-1} by the rule

$$\bar{\pi}(k) = \pi(k) \text{ if } \pi(k) < \pi(n),$$
$$= \pi(k) - 1 \text{ if } \pi(k) > \pi(n).$$

Intuitively, if $\pi(n) = i$, then $\bar{\pi}$ orders the elements $(1, \ldots, n-1)$ in the same way that π orders $(1, \ldots, \hat{i}, \ldots, n)$. (Here and below the notation '$\hat{a}$' denotes the omission of 'a'.) Indeed, the mapping is one - one from $S_{n,i} = \{\pi \in S_n: \pi(n) = i\}$ onto S_{n-1} for each i. Notice that, for $\pi \in S_{n,i}$,

$$(4.2) \qquad \bar{\pi}(x_1, \ldots, \hat{x}_i, \ldots, x_n) = (x_{\pi(1)}, \ldots, x_{\pi(n-1)}).$$

Some additional information about the mapping $\pi \to \bar{\pi}$ is recorded in the following lemma, after which the proof of Theorem 1 will resume.

<u>Lemma 4.1</u>. <u>Let</u> $A \subseteq R^{n-1}$ <u>and let</u> α' <u>be the marginal distribution of</u> α <u>on the first</u> n-1 <u>coordinates</u>. <u>Then, for each</u> $\pi \in S_n$,

$$(4.3) \quad \alpha\{x \in R^n: \pi(x) \in A \times R\} = \alpha'\{y \in R^{n-1}: \bar{\pi}(y) \in A\},$$

<u>and</u>

$$(4.4) \quad \beta(A \times R) = \frac{1}{(n-1)!} \sum_{\pi' \in S_{n-1}} \alpha'\{y \in R^{n-1}: \pi'(y) \in A\}.$$

<u>Proof</u>: Let $\pi \in S_{n,i}$. Then

$$\alpha\{x: \pi(x) \in A \times R\} = \alpha\{x: (x_{\pi(1)}, \ldots, x_{\pi(n-1)}) \in A\}$$
$$= \alpha\{x: (x_1, \ldots, \hat{x}_i, \ldots, x_n) \in \bar{\pi}^{-1}(A)\}$$
$$= \alpha'\{y: \bar{\pi}(y) \in A\}.$$

The second equality in this calculation is by (4.2) and the final one uses condition (v). This proves (4.3).

The calculation below yields (4.4).

$$\beta(A \times R) = \frac{1}{n!} \sum_{\pi \in S_n} \alpha[\pi(x) \in A \times R]$$

$$= \frac{1}{n!} \sum_{\pi \in S_n} \alpha'[\bar{\pi}(y) \in A]$$

$$= \frac{1}{(n-1)!} \sum_{\pi' \in S_{n-1}} \alpha'[\pi'(y) \in A].$$

The successive equalities are by (4.1), (4.3), and the fact that each π' in S_{n-1} is the image of n elements in S_n under the map $\pi \to \bar{\pi}$.

This completes the proof of the lemma. $\square$

To finish the proof of the first assertion of Theorem 1, let $A \subseteq R^{n-1}$ and let $1 \leq i \leq n$. It suffices to verify (2.1).

By condition (iv),

$$\alpha[x_i = x_{(i)}] = 1.$$

Let $C_i = [x_n = x_{(i)}]$. Then

$$\alpha[\pi(x) \in C_i] = 1 \text{ or } 0$$

according as $\pi(n) = i$ or $\pi(n) \neq i$.

Thus, for $B = A \times R$,

$$\beta(C_i \cap B) = \frac{1}{n!} \sum_{\pi \in S_n} \alpha[\pi(x) \in C_i \cap B]$$

$$= \frac{1}{n!} \sum_{\pi \in S_{n,i}} \alpha[\pi(x) \in B]$$

$$= \frac{1}{n!} \sum_{\pi \in S_{n,i}} \alpha'[\bar{\pi}(y) \in A]$$

$$= \frac{1}{n} \left\{\frac{1}{(n-1)!} \sum_{\pi' \in S_{n-1}} \alpha'[\pi'(y) \in A]\right\}$$

$$= \frac{1}{n} \beta(A \times R)$$

One half of the proof of Theorem 1 is now complete. The converse half uses the following two lemmas, the first of which is adopted from an argument in Hill ((1968), p. 688).

Lemma 4.2. Suppose $n > 1$, $\beta \in H_n$, and β' is the marginal distribution of β on the first $n-1$ coordinates. Then $\beta' \in H_{n-1}$.

Proof: The result is clear if $n = 2$; so assume $n = m + 2 > 2$. It remains clear that β' satisfies (i) and (ii). To check (iii), let $J_0, \dots, J_m$ be the intervals into which R is partitioned by the points $x_1, \dots, x_m$ and, for $i = 0, \dots, m$, define

$$U_i = \{x \in R^n: x_{m+2} \in J_i\},$$

$$V_i = \{x \in R^n: x_{m+1} \in J_i\}.$$

For $A \subseteq R^m$, it follows from exchangeability that

$$\beta\{U_i \cap [(x_1, \dots, x_m) \in A]\} = \beta\{V_i \cap [(x_1, \dots, x_m) \in A]\}$$

Denote the common value by ϵ_i, and let $s = \sum_{i=0}^{m} \epsilon_i = \beta[(x_1, \dots, x_m) \in A]$.

Then, for each i,

$$\begin{aligned}
\epsilon_i &= \beta\{U_i \cap [(x_1, \dots, x_m) \in A]\} \\
&= \sum_{j=0}^{m} \beta\{V_j \cap U_i \cap [(x_1, \dots, x_m) \in A]\} \\
&= \sum_{j=0}^{m} [(1 + \delta_{ij})/(m + 2)]\, \epsilon_j \\
&= (s + \epsilon_i)/(m + 2).
\end{aligned}$$

The next to last equality uses condition (iii). It now follows that $\epsilon_i = s/(m + 1) = s/(n-1)$, which shows β' satisfies (iii). □

Set $\mathrm{ord}_i(x) = (x_{(1)}, \dots, \hat{x}_{(i)}, \dots, x_{(n)})$.

Lemma 4.3. Let $n > 1$, $\beta \in H_n$, and let $A \subseteq \{y \in R^{n-1}: y_1 < y_2 < \dots < y_{n-1}\}$. Then for $1 \le i \le n$,

$$(4.5)\quad \beta[\mathrm{ord}_i(x) \in A] = (n-1)!\, \beta[(x_1, \ldots, x_{n-1}) \in A].$$

Proof: Obviously,

$$(4.6)\quad \beta[\mathrm{ord}_i(x) \in A] = \sum_{k=1}^{n} \beta[\mathrm{ord}_i(x) \in A,\ x_n = x_{(k)}].$$

Consider each of the summands on the right. Examine first the summand for $k = i$ and notice that $x_n = x_{(i)}$ implies $\mathrm{ord}_i(x) = \mathrm{ord}(x_1, \ldots, x_{n-1})$. Thus

$$\begin{aligned}(4.7)\quad \beta[\mathrm{ord}_i(x) \in A,\ x_n = x_{(i)}] &= \sum_{\pi \in S_{n-1}} \beta[\pi(x_1, \ldots, x_{n-1}) \in A,\ x_n = x_{(i)}] \\ &= (n-1)!\, \beta[(x_1, \ldots, x_{n-1}) \in A,\ x_n = x_{(i)}] \\ &= \frac{(n-1)!}{n} \beta[(x_1, \ldots, x_{n-1}) \in A].\end{aligned}$$

The second equality is by condition (i) and the final equality is by condition (iii).

Suppose now that $k \neq i$. Then

$$\begin{aligned}(4.8)\quad & \beta[\mathrm{ord}_i(x) \in A,\ x_n = x_{(k)}] \\ &= \sum_{j=1}^{n-1} \beta[\mathrm{ord}(x_1, \ldots, \hat{x}_j, \ldots, x_n) \in A,\ x_j = x_{(i)},\ x_n = x_{(k)}] \\ &= \sum_{j=1}^{n-1} \beta[\mathrm{ord}(x_1, \ldots, x_{n-1}) \in A,\ x_n = x_{(i)},\ x_j = x_{(k)}] \\ &= \beta[\mathrm{ord}(x_1, \ldots, x_{n-1}) \in A,\ x_n = x_{(i)}] \\ &= \beta[\mathrm{ord}_i(x) \in A,\ x_n = x_{(i)}].\end{aligned}$$

The second equality follows from condition (i) when j and n are interchanged.

The result (4.5) now follows from (4.6), (4.7), and (4.8). □

To complete the proof of Theorem 1, let $\beta \in H_n$ and let α be the β-distribution of ord x. That α satisfies (iv) is immediate from condition (ii). Also, (v) is a consequence of Lemma 4.3 in the special case when $m = n-1$. For general m, an inductive argument can be based on Lemma 4.2.

One additional lemma helps to set the stage for the next section. Let

$\beta \in H_n$ and define, for $t \in R$,

$$F(t) = \beta\{x \in R^n: x_1 \le t\}.$$

Lemma 4.4. For every $t \in R$ and, for $i = 1, \ldots, n$,

$$F(t) = \beta[x_i \le t] = \beta[x_{(i)} \le t].$$

Proof: By Theorem 1, the β-distribution of ord x is in O_n. Thus, by condition (v) with $m = 1$, $\beta[x_{(i)} \le t]$ is the same for all $i = 1, \ldots, n$. Furthermore,

$$\beta[x_{(1)} \le t] \le \beta[x_i \le t] \le \beta[x_{(n)} \le t],$$

for all i. Hence, all these probabilities are equal. □

Let

$$L_t = \{x \in R^n: x_i > t, i = 1, \ldots, n\},$$

$$S_t = \{x \in R^n: x_i \le t, i = 1, \ldots, n\}.$$

Corollary. For $\beta \in H_n$ and every $t \in R$, $\beta(L_t) + \beta(S_t) = 1$.

Proof: $\beta(L_t) + \beta(S_t) = \beta[x_{(1)} > t] + \beta[x_{(n)} \le t] = 1 - F(t) + F(t)$. □

5. A solution for arbitrary finite n.

Let ν be a probability R and define, for each positive integer n, the strategic product measure ν^n on R^n by the formula

$$(5.1) \quad \nu^n(A) = \int\ldots\int A(x_1, \ldots, x_n)\, \nu(dx_n)\ldots\nu(dx_1)$$

for $A \subseteq R^n$. In (5.1), the set A has been identified with its indicator function. This useful convention, which is due to de Finetti, is followed below.

If ν were countably additive, then ν^n would be the usual product measure and the order of integration in (5.1) would be irrelevant. Here the order can be crucial, as will be seen in Lemma 5.3 below.

The symmetrization $\rho(\nu^n)$ of ν^n is called a symmetric product measure because it is clearly exchangeable and is also a product measure in the sense that it agrees with ν^n on sets of the form $A_1 \times \ldots \times A_n$,

$A_i \subseteq R$. For countably additive ν, ν^n and $\mathcal{P}(\nu^n)$ agree on the Borel sets as well.

If $t \in R$ and if, for every $\epsilon > 0$,

(5.2) $\quad \nu(t, t+\epsilon) = 1 \quad (\nu(t-\epsilon, t) = 1)$

then ν is concentrated at $t+$ $(t-)$. Similarly, if

(5.3) $\quad \nu(a, +\infty) = 1 \ (0)$

for all a, then ν is concentrated at $+\infty$ $(-\infty)$.

Theorem 2. Let $n \geq 2$ and let ν be a probability on R. Then $\mathcal{P}(\nu^n) \in H_n$ if, and only if, ν is concentrated at $+\infty$ or $-\infty$, or at $t+$ or $t-$ for some $t \in R$.

Proof: Let $\mu_n = \mathcal{P}(\nu^n)$. The following two lemmas are helpful in the proof of necessity.

Lemma 5.1. If $\mu_n \in H_n$, then $\nu\{x\} = 0$ for every $x \in R$.

Proof: If the conclusion were false, then, as is easy to see, μ_n would violate condition (ii) (no ties). □

Let $F(x) = \nu(-\infty, x]$ for $x \in R$.

Lemma 5.2. If $\mu_n \in H_n$, then for every $x \in R$, $F(x) = 0$ or $F(x) = 1$.

Proof: Suppose, by way of contradiction, that $F(t) = \lambda$ for some $t \in R$, $0 < \lambda < 1$. Let

$$L = \{x: x_1 > t, \ldots, x_n > t\},$$

$$S = \{x: x_1 \leq t, \ldots, x_n \leq t\}.$$

Both L and S are invariant under permutations of the coordinates and, hence,

$$\mu_n(L) = \nu^n(L) \text{ and } \mu_n(S) = \nu^n(S).$$

By (5.1), $\nu^n(L) = (1-\lambda)^n$ and $\nu^n(S) = \lambda^n$. Therefore,

$$\begin{aligned} \mu_n(L) + \mu_n(S) &= \lambda^n + (1-\lambda)^n \\ &< \lambda + (1-\lambda) \\ &= 1, \end{aligned}$$

which contradicts the corollary to Lemma 4.4. □

To prove the direct implication of Theorem 2, suppose $\mu_n \in H_n$. If $F(x) = 0$ (1) for all $x \in R$, then, as is easily seen, ν is concentrated at $-\infty$ $(+\infty)$. If F assumes both values 0 and 1, let

$$t = \inf\{x: F(x) = 1\}.$$

If $F(t) = 1$ (0), then ν is concentrated at $t-$ $(t+)$.

Lemma 5.3. (i) If ν is concentrated at $+\infty$ or at $t-$, then

$$\nu^n\{x \in R^n: x_1 < x_2 < \ldots < x_n\} = 1.$$

(ii) If ν is concentrated at $-\infty$ or $t+$, then

$$\nu^n\{x \in R^n: x_n < x_{n-1} < \ldots < x_1\} = 1.$$

Proof: The proof will be given for a ν concentrated at $t-$. (The other cases can be handled similarly or reduced to this one by appropriate transformations.)

Let $A_1 = (-\infty, t)$ and, for $n \geq 2$, let

$$A_n = \{x \in R^n: x_1 < x_2 < \ldots < x_n < t\}.$$

Then, for $n \geq 2$,

$$\int A_n(x_1, \ldots, x_{n-1}, x_n)\nu(dx_n) = A_{n-1}(x_1, \ldots, x_{n-1}),$$

as follows from (5.2). Now use (5.1) to conclude

$$\nu^n(A_n) = \nu^{n-1}(A_{n-1}) = \ldots = \nu(A_1) = 1. \quad \square$$

The next lemma holds for an arbitrary probability ν on R.

Lemma 5.4. For $1 \leq m < n$, the marginal ν^n distribution on any set of m coordinates (in ascending order) is ν^m.

Proof: It is enough to give the proof for $m = n-1$, for then the general case will follow by an inductive argument. What must be shown is that

$$\nu^n(A) = \nu^{n-1}(B)$$

for $B \subseteq R^{n-1}$ and $A = \{x \in R^n: (x_1, \ldots, \hat{x}_i, \ldots, x_n) \in B\}$. Fix $1 < i < n$. Then

$$\nu^n(A) = \int\ldots\int A(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)\nu(dx_n)\ldots\nu(dx_1)$$

$$= \iiint A(u, x_i, v)\nu^{n-i}(dv)\nu(dx_i)\nu^{i-1}(du)$$

$$= \iiint B(u,v)\nu^{n-i}(dv)\nu(dx_i)\nu^{i-1}(du)$$

$$= \iint B(u,v)\nu^{n-i}(dv)\nu^{i-1}(du)$$

$$= \int\cdots\int B(x_1,\ldots,x_{n-1})\nu(dx_{n-1})\cdots\nu(dx_1)$$

$$= \nu^{n-1}(B).$$

For $i = 1$ or $i = n$, the calculation is similar but even simpler. $\square$

It is now easy to prove the converse half of Theorem 2. Indeed, if ν is concentrated at $+\infty$ or $-t$, then, by Lemmas 5.3 and 5.4, $\nu^n \in O_n$ and, hence, by Theorem 1, $\mathcal{P}(\nu^n) \in H_n$. If ν is concentrated at $-\infty$ or $t+$, let α be the ν^n-distribution of $r(x)$, where $r(x_1,\ldots,x_n) = (x_n,\ldots,x_1)$. Then, by Lemmas 5.3 and 5.4, $\alpha \in \dot{O}_n$ and, by Proposition 1, $\mathcal{P}(\alpha) \in H_n$. But $\bar{\mathcal{P}}(\alpha) = \mathcal{P}(\nu^n)$. This completes the proof of Theorem 2.

6. Extension to the infinite case.

Consider now the question of the existence of an infinite sequence of variables $X_1, X_2 \ldots$ such that, for every n, $X_1,\ldots,X_n$ satisfy Hill's conditions. In view of the preceding results, it is not difficult to see that such sequences exist and this is recorded in the next theorem.

Theorem 3. *There is a finitely additive probability measure μ_∞ defined on the subsets of R^∞ whose marginal on the first n coordinates is in H_n for every n.*

Proof: The notation of section 5 will be used. Thus ν is a probability on R, ν^n is the probability on R^n defined by (5.1), and $\mu_n = \mathcal{P}(\nu^n)$. Assume that ν is concentrated at $t+$, $t-$, $+\infty$, or $-\infty$, so that Theorem 2 implies $\mu_n \in H_n$ for every n. The following lemma shows that the μ_n are consistent with a probability on R^∞.

<u>Lemma 6.1. For $1 \le m < n$, the μ_n marginal distribution on any set of m coordinates is μ_m.</u>

<u>Proof</u>: By induction, it is enough to treat the case $m = n-1$ and, by the exchangeability of μ_n, there is no loss of generality in taking the first m coordinates.

Let $A \subseteq R^{n-1}$ and calculate

$$\begin{aligned}\mu_n(A \times R) &= \frac{1}{n!} \sum_{\pi \in S_n} \nu^n[\pi(x) \in A \times R] \\ &= \frac{1}{n!} \sum_{\pi \in S_n} \nu^{n-1}[\overline{\pi}(y) \in A] \\ &= \frac{1}{(n-1)!} \sum_{\pi' \in S_{n-1}} \nu^{n-1}[\pi'(y) \in A] \\ &= \mu_{n-1}(A).\end{aligned}$$

Here the first and last equalities are by definition of μ_n and μ_{n-1}, the second is by Lemmas 4.1 and 5.4, and the third holds because each π' in S_{n-1} is the image of n elements in S_n under the mapping $\pi \to \overline{\pi}$, which is defined just before Lemma 4.1. □

Return to the proof of Theorem 3 and define

$$\mu_\infty(B \times R^\infty) = \mu_n(B)$$

for $B \subseteq R^n$. By Lemma 6.1, this definition makes sense. Although the Kolmogorov extension does not apply in the present finitely additive setting, it is possible to extend μ_∞ to all subsets of R^∞ using a finitely additive technique such as the Hahn-Banach theorem. □

There is an alternative approach to the proof of Theorem 3 in which the definition of μ_∞ imitates that of μ_n on R^n. First, let ν^∞ be the distribution on R^∞ such that, under ν^∞, x_1 has distribution ν and, for every $n \ge 1$, given $x_1,\ldots,x_n$, the conditional distribution of x_{n+1} is

ν. The measure ν^∞ is strategic in the sense of Dubins and Savage (1965) and has a natural extension to the Borel subsets of R^∞ described by Purves and Sudderth (1976). The μ_∞ distribution is obtained by choosing a point at random according to ν^∞ and then putting the coordinates in a random order. To make the second step precise, let S_∞ be the group of those permutations of the natural numbers which leave all but a finite number of coordinates fixed. The group S_∞ has an invariant, finitely additive probability γ as is perhaps well-known and not too difficult to prove. The measure γ can be used to choose the random order of the coordinates, and μ_∞ can be defined by the following formula

$$\mu_\infty(A) = \int \nu^\infty\{x\colon \ \pi(x) \in A\}\ \gamma(d\pi),$$

where A is a Borel subset of R^∞, $\pi \in S_\infty$, $x = (x_1, x_2, \ldots) \in R^\infty$, and $\pi(x) = (x_{\pi(1)}, x_{\pi(2)}, \ldots)$.

7. Extreme points of H_n.

Are there members of H_n different from the symmetric product measures $\rho(\nu^n)$ of Theorem 2? The answer is 'yes' because H_n is a convex set of measures and, hence, mixtures of elements in H_n are again in the set. The next question is whether there are members of H_n which are not mixtures of the symmetric product measures in H_n. The answer remains 'yes' as the following example demonstrates for H_2.

Example. Let π be a translation invariant probability on R and let β be that probability on R^2 such that the β-distribution of x_1 is π and such that the β-conditional distribution of x_2 given x_1 assigns probability $\frac{1}{2}$ to each of the points $x_1 + 1$ and $x_1 - 1$. Formally, for $A \subseteq R^2$

$$\beta(A) = \tfrac{1}{2}\,\pi\{x\colon \ (x, x+1) \in A\} + \tfrac{1}{2}\,\pi\{x\colon \ (x, x-1) \in A\}.$$

The measure β is in H_2 and is not a mixture of the symmetric product measures in H_2.

To check that β is in H_2 is easy. Conditions (i) and (iii) are immediate, and the following calculation verifies condition (ii):

If $r(A) = \{(x, y): (y, x) \in A\}$, then

$$\begin{aligned}\beta(r(A)) &= \tfrac{1}{2}\,\pi(x: (x, x+1) \in r(A)) + \tfrac{1}{2}\,\pi(x: (x, x-1) \in r(A)) \\ &= \tfrac{1}{2}\,\pi(x: (x+1, x) \in A) + \tfrac{1}{2}\,\pi(x: (x-1, x) \in A) \\ &= \tfrac{1}{2}\,\pi(x: (x, x-1) \in A) + \tfrac{1}{2}\,\pi(x: (x, x+1) \in A) \\ &= \beta(A).\end{aligned}$$

All but the third equality are definitions; the third equality follows from the translation invariance of π.

To see that β cannot be represented as a mixture of symmetric product measures, note that $\beta\{(x_1,x_2): |x_1 - x_2| = 1\} = 1$ while for any symmetric product measure γ in H_2--and hence any mixture of them--Theorem 2 implies that $\gamma\ \{(x_1,x_2): \epsilon < |x_1 - x_2| < M\} = 0$ for all positive ϵ and M.

The measure β has the same properties if the conditional distribution of x_2 given x_1 is taken to be that of $Y + x_1$ where Y has any countably additive distribution which is symmetric about zero. The proof that β is in H_2 uses Theorem 3 of Heath and Sudderth (1976).

Despite the example, it is true that every measure in H_n is a mixture of probabilities which have much in common with the symmetric product measures. To see this, view H_n as being a subspace of the collection F of all functions from subsets of R^n to $[0,1]$. Equip F with the topology of pointwise convergence under which it is compact. Then H_n is a convex, closed (and, hence, compact) subset of F, as is easy to verify. By the Krein-Milman Theorem (Dunford and Schwartz (1957), Theorem V.8.4), H_n is the closed, convex hull of its extreme points or, what amounts to the same thing, every measure in H_n is a (finitely additive) mixture of the extreme points. We have not been able to characterize the set of extreme points of H_n, but some information about them is given below.

Hewitt and Savage (1955) did characterize the extreme points of the exchangeable measures on R^n and R^∞. However, they restricted their study of finitely additive measures to those defined only on cylinder sets. Such a simplifying restriction is not possible here because the event $[x_n = x_{(i)}]$, which occurs in the definition of H_n, is not a cylinder set.

A probability β on R^n is <u>concentrated at</u> t+ (t-) <u>on the diagonal</u> if, for every $\epsilon > 0$,

$$\beta\{x \in R^n: t < x_i < t + \epsilon,\ i = 1, \ldots, n\} = 1$$

$$(\beta\{x \in R^n: t - \epsilon < x_i < t,\ i = 1, \ldots, n\} = 1).$$

Similarly, β is <u>concentrated at</u> $+\infty$ $(-\infty)$ if, for every $a \in R$,

$$\beta\{x \in R^n: x_i > a,\ i = 1, \ldots, n\} = 1$$

$$(\beta\{x \in R^n: x_i < a,\ i = 1, \ldots, n\} = 1).$$

Let C_n be the collection of those β which are concentrated in one of the senses above. The measures $\mathcal{P}(\nu^n)$, which occur in Theorem 2, are easily seen to be elements of C_n. Let $\mathcal{E}_n$ be the set of extreme points of H_n.

<u>Theorem 4</u>. For $n \geq 2$, $\mathcal{E}_n \subseteq C_n$.

<u>Proof</u>: Let $\beta \in \mathcal{E}_n$ and $F(t) = \beta[x_1 \leq t]$ for $t \in R$. If, for some $t \in R$ and $\lambda \in (0,1)$, $F(t) = \lambda$, then $\beta = \lambda\beta_S + (1-\lambda)\beta_L$ where β_S and β_L are probabilities on R^n defined by

$$\beta_S(A) = \lambda^{-1}\beta([x_1 \leq t] \cap A),$$

$$\beta_L(A) = (1-\lambda)^{-1}\beta([x_1 > t] \cap A].$$

With the aid of Lemma 4.4, it is not difficult to check that β_S and β_L are in H_n. But this contradicts our assumption that β is extreme. Hence, no such λ can exist and $F(t) = 0$ or 1 for all t.

If $F(t) = 0$ (1) for all t, then β is concentrated at $+\infty$ $(-\infty)$ as Lemma 4.4 implies. If F assumes both the values 0 and 1 and $t = \inf\{s: F(s) = 1\}$, then β is concentrated at t+ or t- on the diagonal according as $F(t) = 0$ or 1 as follows again from Lemma 4.4. ☐

From Theorem 4 and the Krein-Milman Theorem, it follows that, for $n \geq 2$, every measure in H_n is a mixture of concentrated measures.

8. The distribution of the range.

The *range* of a vector $x \in R^n$ is defined to be $r = r(x) = x_{(n)} - x_{(1)}$. Further information about the measures in H_n is obtained in this section in terms of the distribution of the range r. The first result, which is almost a corollary of Theorem 4, is that, for $n \geq 2$, if the values of x cannot be arbitrarily large, then r must be concentrated near zero.

Theorem 5. Let $n \geq 2$, $\beta \in H_n$ *and suppose* β *has compact support. Then* $\beta[r < \epsilon] = 1$ *for every* $\epsilon > 0$

Proof: Since β has compact support, there is a finite interval $[a,b]$ such that $\beta[a \leq x_{(1)} \leq x_{(n)} \leq b] = 1$. For $\epsilon > 0$, let $a = t_0 < t_1 < \dots < t_n = b$ be a partition of $[a,b]$ such that $t_{i+1} - t_i < \epsilon$ for $i = 0, \dots, n-1$. Then

$$[r > \epsilon] \subseteq \bigcup_{k=0}^{n} [x_{(1)} \leq t_k < x_{(n)}].$$

Furthermore, for every k,

$$\beta[x_{(1)} \leq t_k < x_{(n)}] = \beta[x_{(1)} \leq t_k] - \beta[x_{(n)} \leq t_k] = 0,$$

by Lemma 4.4. □

If $n = 2$, then the example of section 7 shows that r can have any countably additive distribution when β does not have compact support. However, for $n \geq 3$, the distribution of r must be concentrated near 0 and ∞, as the following theorem shows.

Theorem 6. Let $n \geq 3$ *and* $\beta \in H_n$. *Then* $\beta\{x \in R^n: \epsilon < r(x) < M\} = 0$ *for all positive numbers* ϵ *and* M.

Proof: The proof is based on the equality

$$r = d_2 + \cdots + d_n, \tag{8.1}$$

where $d_i = x_{(i)} - x_{(i-1)}$, and on the fact that each of the nonnegative functions $r, d_2, \dots, d_n$ have the same distribution under β. This latter fact is because the distribution of ord x under β is in O_n (Theorem 1) so that condition (v) in the definition of O_n applies.

Let $t > 0$. By (8.1), the event $[r > t]$ contains $[d_i > t]$ for each i. However, these two events have the same measure under β. Hence, they can differ only by a set of β-measure zero. Consequently, the events $[r > t]$ and $[d_2 > t, \ldots, d_n > t]$ also differ only by a β-null set. Now calculate as follows:

$$\begin{aligned}\beta[r > t] &= \beta[d_2 > t, \ldots, d_n > t] \\ &\le \beta[d_2 + \cdots + d_n > (n-1)t] \\ &= \beta[r > (n-1)t].\end{aligned}$$

It follows that $\beta[t < r \le (n-1)t] = 0$ and, more generally, that

(8.2) $\quad \beta[t_1 < r < t_2] = 0$

for $0 < t_1 < t_2 < \infty$.

<u>Corollary</u>. <u>Let</u> $n \ge 3$ <u>and</u> $\beta \in H_n$. <u>Then there exist</u> β_0 <u>and</u> $\beta_\infty \in H_n$ <u>and</u> $\lambda \in [0,1]$ <u>such that</u>

(8.3) $\quad \beta = \lambda\beta_0 + (1-\lambda)\beta_\infty$

<u>and such that, for every</u> $\epsilon > 0$,

(8.4) $\quad \beta_0[r < \epsilon] = \beta_\infty[r > \epsilon] = 1.$

<u>Proof</u>: To define the two components β_0, β_∞, first fix $t > 0$ and let $\lambda = \beta[r \le t]$. If $\lambda = 1$ $(\lambda = 0)$, set β_0 $(\beta_\infty) = \beta$ and let β_∞ (β_0) be any measure in H_n which satisfies (8.4). If $0 < \lambda < 1$, define

$$\beta_0(A) = \lambda^{-1}\beta(A \cap [r \le t]),$$
$$\beta_\infty(A) = (1-\lambda)^{-1}\beta(A \cap [r > t])$$

for $A \subseteq R^n$. Clearly, (8.3) holds and (8.4) follows from (8.2). It remains only to check that β_0, $\beta_\infty \in H_n$. The conditions of exchangeability and no ties are trivial to verify. The proof of condition (iii) requires a preliminary observation.

$$\begin{aligned}[r > t] &\supset [|x_2 - x_1| > t] \\ &\supset [d_2 > t, \ldots, d_n > t].\end{aligned}$$

As previously noted, the first and last events differ by a β-null set.

Thus the same must be true of the first and second events. Now let $A = B \times R$ for some $B \subseteq R^{n-1}$ and let $C_i = [x_n = x_{(i)}]$. Then

$$\begin{aligned}
\beta_\infty(A \cap C_i) &= (1-\lambda)^{-1}\beta(A \cap [r > t] \cap C_i) \\
&= (1-\lambda)^{-1}\beta(A \cap [|x_2 - x_1| > t] \cap C_i) \\
&= (1-\lambda)^{-1}n^{-1}\beta(A \cap [|x_2 - x_1| > t]) \\
&= (1-\lambda)^{-1}n^{-1}\beta(A \cap [r > t]) \\
&= n^{-1}\beta_\infty(A).
\end{aligned}$$

The third equality in the calculation uses condition (iii) for β. This verifies (iii) for β_∞. A similar calculation does the same for β_0. $\square$

9. A formula for predictive distributions.

Consider a typical statistical framework in which $X_1, X_2, \ldots, X_{n+1}$ are independent random variables each having the (countably additive) distribution $\alpha(\theta)$. Suppose θ has a (prior) distribution μ and that $\nu(x_1, \ldots, x_n)$ is a (posterior) distribution for θ given $X_1 = x_1, \ldots, X_n = x_n$. Then a (predictive) distribution for X_{n+1} given $X_1 = x_1, \ldots, X_n = x_n$ is $\gamma(x_1,\ldots,x_n)$ where

(9.1) $$\gamma(x_1, \ldots, x_n)(A) = \int\alpha(\theta)(A)\ \nu(x_1,\ldots,x_n)(d\theta),$$

for A a Borel subset of R. This is a standard formula in the conventional, countably additive theory. The object of this section is to verify it when the prior distribution μ is only finitely additive and, in particular, to verify (3.1) which is a special case of (9.1). The proof is based on two lemmas whose statements require the following definition.

For a nonempty set Y, let $\mathfrak{m}(Y)$ be the set of finitely additive probabilities on Y. Consider nonempty sets $Y_1,\ldots,Y_{n+1}$ and let $Y^k = Y_1 \times \ldots \times Y_k$, $1 \le k \le n+1$. A <u>strategy</u> σ on Y^{n+1} is a sequence $\sigma_0,\sigma_1,\ldots,\sigma_n$ where $\sigma_0 \in \mathfrak{m}(Y_1)$ and, for $1 \le k \le n$, σ_k is a mapping from Y^k to $\mathfrak{m}(Y_{k+1})$. Each strategy σ on Y^{n+1} determines a probability, also denoted σ, in $\mathfrak{m}(Y^{n+1})$, which is defined by the formula

(9.2) $\sigma g = \int\int \ldots \int g(y_1, \ldots, y_{n+1})\sigma_n(y_1, \ldots, y_n)(dy_{n+1}) \ldots \sigma_1(y_1)(dy_2)\sigma_0(dy_1)$

for bounded functions g from Y^{n+1} to R. A probability in $\mathfrak{M}(Y^{n+1})$ is called <u>strategic</u> if it arises from a strategy in this manner. Roughly speaking, strategic probabilities are those which can be defined by a system of conditional distributions. As is well-known, countably additive probabilities on sufficiently regular product spaces can always be defined via conditionals. However, there do exist finitely additive probabilities which are far from strategic (see Dubins (1975)). All mention of sigma-fields and measurability is suppressed in the rest of this section. However, it is easy to see that the conclusions generalize to a situation in which each Y_i is equipped with a sigma-field and only product measurable functions on Y^{n+1} are considered.

<u>Lemma 9.1. Let σ be a strategy on $Y_1 \times Y_2 \times Y_3$ and let σ' be the marginal distribution of σ on $Y_1 \times Y_3$. Then σ' is also strategic with $\sigma_0' = \sigma_0$ and</u>

(9.3) $\sigma_1'(y_1)\varphi = \int\int \varphi(y_3)\sigma_2(y_1,y_2)(dy_3)\sigma_1(y_1)(dy_2)$

<u>for φ a bounded function from Y_3 to R.</u>

<u>Proof</u>: Let g' be a bounded function from $Y_1 \times Y_3$ to R and set

$$g(y_1,y_2,y_3) = g'(y_1, y_3).$$

Then

$$\begin{aligned}\sigma' g' &= \sigma g \\ &= \int\int\int g'(y_1,y_3)\sigma_2(y_1,y_2)(dy_3)\sigma_1(y_1)(dy_2)\sigma_0(dy_1) \\ &= \int\int g'(y_1,y_3)\sigma_1'(y_1)(dy_3)\sigma_0(dy_1).\end{aligned}$$

The first equality is by definition of the marginal distribution, the second is by (9.2), and the third by (9.3). $\square$

<u>Lemma 9.2. Let σ be a strategy on $Y_1 \times Y_2 \times Y_3$ and let β be the measure induced by σ on $Y_2 \times Y_1 \times Y_3$ by reversing the first two coordinates. If the marginal distribution of β on $Y_2 \times Y_1$ is strategic,</u>

<u>then</u> β <u>is strategic on</u> $Y_2 \times Y_1 \times Y_3$ <u>and</u> $\beta_2(y_2,y_1)$ <u>can be taken to be</u> $\sigma_2(y_1,y_2)$.

<u>Proof</u>: Let β' be the marginal of β on $Y_2 \times Y_1$. Because β' is induced by a strategy β_0,β_1, it follows that, for bounded functions g' from $Y_2 \times Y_1$ to R,

$$\begin{aligned}\int g'd\beta' &= \iint g'(y_2,y_1)\beta_1(y_2)(dy_1)\beta_0(dy_2)\\ &= \iint g'(y_2,y_1)\sigma_1(y_1)(dy_2)\sigma_0(dy_1).\end{aligned}$$

Thus, if g is a bounded function from $Y_2 \times Y_1 \times Y_3$ to R,

$$\begin{aligned}&\iiint g(y_2,y_1,y_3)\sigma_2(y_1,y_2)(dy_3)\beta_1(y_2)(dy_1)\beta_0(dy_2)\\ &\qquad = \iiint g(y_2,y_1,y_3)\sigma_2(y_1,y_2)(dy_3)\sigma_1(y_1)(dy_2)\sigma_0(dy_1)\\ &\qquad = \int g\, d\beta.\end{aligned}$$

The desired conclusions now follow. □

To derive (9.1) from the lemmas, let $Y_1 = \Theta$, $Y_2 = R^n$, and $Y_3 = R$. Let σ be the strategy on $Y_1 \times Y_2 \times Y_3$ such that $\sigma_0 = \mu$ (prior on Θ), $\sigma_1(\theta)$ is the product measure $\alpha(\theta)^n$ on R^n, and $\sigma_2(\theta,y_2) = \alpha(\theta)$. Next let β be the measure obtained from σ by reversing Θ and Y_2 as in Lemma 9.2. By assumption, there is a (posterior) distribution for θ given y_2. Thus the marginal of β on $Y_2 \times \Theta$ is strategic. Hence, by Lemma 9.2, β is strategic and

$$\beta_2(y_2,\theta) = \sigma_2(\theta,y_2) = \alpha(\theta)$$

Now apply Lemma 9.1 to see that the marginal β' of β on $Y_2 \times Y_3$ is strategic with

$$\beta'(y_2)\varphi = \iint \varphi(y_3)\alpha(\theta)(dy_3)\beta_1(y_2)(d\theta)$$

for bounded functions φ from Y_3 to R. This final formula is the same, except for notation, as (9.1).

10. Two questions.

The natural way to attempt a construction of a measure in H_n is to specify its successive conditional distributions. One selects a distribution for X_1 and then a conditional distribution for X_2 given X_1 in such a way that the distribution of (X_1, X_2) is an element of H_2. Such a construction is presented in the example of section 7. However, we have not been able to continue in this fashion and specify a distribution for X_3 given (X_1, X_2) so that the joint distribution of (X_1, X_2, X_3) lies in H_3. We do not know whether this can be done. Equivalently, we do not know whether there exist any strategic measures in H_n for $n \geq 3$.

According to a famous theorem of de Finetti, every infinite sequence of exchangeable variables is a mixture of sequences of independent variables. A very general version of the theorem was proved by Hewitt and Savage (1955). However, their results do not settle the question of whether any (or all) of the exchangeable measures on R^∞ of section 6 can be represented as a finitely additive mixture of countably additive, independent, identically distributed variables. If such a representation were possible, the mixing measure would be a natural noninformative prior on the space of countably additive distribution functions. Recall that Jeffrey's prior on normal distributions did lead (in section 3) to an element of H_3.

REFERENCES

Dubins, Lester and Leonard J. Savage (1965), Inequalities for Stochastic Processes: How to Gamble if You Must, Dover edition (1976).

Dubins, Lester E. (1975), "Finitely additive conditional probabilities, conglomerability, and disintegrations," Ann. Probability, 3 p. 89-99.

Dunford, Nelson and Jacob Schwartz (1957), Linear Operators Part I, Interscience.

Greenleaf, Frederich P. (1969), Invariant Means on Topological Groups, Van Nostrand.

Heath, David and William D. Sudderth (1976), "On Finitely Additive Priors, Coherence, and Extended Admissibility," University of Minnesota Technical Report No. 278, to appear Annals of Statistics.

Hewitt, Edwin and Leonard J. Savage (1955), "Symmetric Measures on Cartesian Products," Trans. AMS, Vol. 80, pp. 470-501.

Hill, Bruce (1968), "Posterior Distribution of Percentiles: Bayes Theorem for Sampling from a Population," JASA, Vol.63, pp. 677-691.

Jeffreys, Harold (1932), "On the Theory of Errors and Least Squares," Proceedings of the Royal Society of London Series A Vol. 138, pp. 48-54.

Purves, Roger and William D. Sudderth, (1976), "Some Finitely Additive Probability," Annals of Probability, Vol. 4, pp. 259-276.